

EXHIBIT 103

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Nikolay Bashlykov 30(b)(6)
Highly Confidential - Attorneys' Eyes Only

Page 1

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

RICHARD KADREY, et al.,)	
Individual and)	
Representative)	
Plaintiffs,)	
)	
v.)	Case No.:
)	3:23-cv-03417-VC
META PLATFORMS, INC.,)	
Defendant.)	

** HIGHLY CONFIDENTIAL - ATTORNEYS' EYES ONLY **

Videotaped 30(b)(6) deposition of Defendant
META PLATFORMS, INC.,
by and through its corporate designee
NIKOLAY BASHLYKOV
Friday, December 6, 2024

London, England
United Kingdom

Reported stenographically by:
Leah M. Willersdorf,
RMR, CRR, FBIVR, ACR, QRR2*, CLR

DIGITAL EVIDENCE GROUP
1730 M. Street, NW, Suite 812
Washington, D.C. 20036
(202) 232-0646

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 16

1 A. Actually, if you could remind me. 12:40:35

2 Q. So the testimony you gave yesterday and 12:40:37

3 today, do you understand that testimony to be 12:40:41

4 testimony you've given in your personal capacity in 12:40:44

5 this case? 12:40:46

6 A. Yes. 12:40:47

7 Q. And how do you understand your testimony 12:40:48

8 in the context of this deposition right now? 12:40:51

9 A. I think it relates to the company. 12:40:56

10 Q. And it is on behalf of the company; do you 12:40:58

11 understand that testimony -- 12:41:00

12 A. Yes. 12:41:01

13 Q. -- to be binding on the company? 12:41:02

14 A. Correct, yes. 12:41:05

15 Q. Okay. 12:41:06

16 We'd like to start with authenticating the 12:41:07

17 hard drives that are in Topic 7 for which you 12:41:10

18 have been offered to provide testimony. 12:41:15

19 Meta_Kadrey_Data_001, 002, 003, 004, and 005, were 12:41:21

20 these -- 12:41:33

21 A. Okay. Should I -- sorry, should I look 12:41:34

22 somewhere? 12:41:37

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 17

1	Q.	No. These are the hard drives that are	12:41:37
2		listed in Topic 7 --	12:41:39
3	A.	Okay.	12:41:41
4	Q.	-- which you can see in the deposition	12:41:41
5		notice for Topic 7. Is data on these hard drives, was	12:41:47
6		it obtained from Meta's files?	12:41:51
7	A.	Yes.	12:41:52
8	Q.	And was this data made in the ordinary	12:41:52
9		course of business at Meta?	12:41:57
10	A.	Yes.	12:41:58
11	Q.	You have also been offered as the	12:41:58
12		corporate deponent on behalf of Meta with respect to	12:42:03
13		Topic 10, which concerns source code; is that correct?	12:42:05
14	A.	Correct.	12:42:10
15	Q.	And to authenticate the source code, which	12:42:10
16		is not here in Britain, all of that source code	12:42:14
17		has been made available in Cooley's Palo Alto office.	12:42:18
18		Is that source code also obtained from	12:42:23
19		Meta's files?	12:42:26
20	A.	Yes.	12:42:27
21	Q.	And was that source code made in the	12:42:27
22		ordinary course of business at Meta?	12:42:30

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 67

1	pretraining Meta LLaMA models?	13:59:30
2	A. It was.	13:59:31
3	Q. Has Books3 Gutenberg been used in	13:59:32
4	post-training of Meta LLaMA models?	13:59:36
5	A. Parts of it was, yes.	13:59:38
6	Q. Do you know which parts?	13:59:41
7	A. It's hard to say, and you can't say	13:59:49
8	document by document, so subset of Books3 was used.	13:59:57
9	Q. Was Books3 Gutenberg, B3G, used for	14:00:01
10	ablation studies?	14:00:07
11	A. For some of the ablation studies, parts of	14:00:11
12	Books3G documents were used.	14:00:19
13	Q. And were those ablation studies in the	14:00:20
14	pretraining part of the life cycle?	14:00:23
15	A. Yes.	14:00:35
16	Q. And were they the ablation studies in the	14:00:35
17	post-training part of the life cycle?	14:00:39
18	A. Yes.	14:00:41
19	Q. And was B3G used for benchmarking? Is	14:00:42
20	that the same as ablations?	14:00:48
21	A. Could you clarify on the benchmarking what	14:00:56
22	do you mean?	14:00:58

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 68

1 Q. How do you define "benchmarking"? It's a 14:00:59

2 term in the documents frequently. 14:01:03

3 A. Mmm. So it depends on the context, but 14:01:08

4 most -- I would refer it as measuring performance of 14:01:11

5 the model against certain benchmarks. 14:01:16

6 Q. Is that like an ablation study -- 14:01:20

7 A. So ablation study, part of the ablation 14:01:24

8 study is measuring the performance on certain 14:01:31

9 benchmarks, like, for specific ablations. 14:01:34

10 Q. Okay. So ablation studies and 14:01:37

11 benchmarking are not always synonymous? 14:01:45

12 A. They are not. 14:01:45

13 Q. Okay. 14:01:46

14 Was B3G used to test memorization? 14:01:52

15 MR. WEINSTEIN: Object to form, outside 14:01:56

16 the scope. 14:01:57

17 MS. POUEYMIROU: No. Topic 7 that you've 14:01:57

18 agreed to is the uses of these datasets. 14:02:00

19 MR. WEINSTEIN: In training the LLaMA 14:02:02

20 models. 14:02:03

21 MS. POUEYMIROU: Memorization is a part of 14:02:05

22 training. 14:02:07

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 69

1	MR. WEINSTEIN: I'm not sure I agree with	14:02:07
2	that. Again, not for here. I'll put the objection on	14:02:08
3	the record. The witness can answer to the best of his	14:02:12
4	knowledge.	14:02:15
5	BY MS. POUEYMIROU:	14:02:15
6	Q. Let me actually ask: Is -- well, we'll	14:02:16
7	stick with that question. Has B3G been used for	14:02:23
8	testing memorization?	14:02:27
9	MR. WEINSTEIN: Same objection; outside	14:02:29
10	the scope.	14:02:40
11	THE WITNESS: To the best of my knowledge,	14:02:40
12	some parts or documents of B3G could be used in	14:02:42
13	memorization, like the project which checks.	14:02:58
14	I wouldn't call that memorization and, probably,	14:03:02
15	I don't know, like, the full term that you -- that	14:03:04
16	copies it, but...	14:03:12
17	BY MS. POUEYMIROU:	14:03:12
18	Q. Like the project which checks for	14:03:12
19	memorization; is that what you meant?	14:03:14
20	A. So the --	14:03:16
21	MR. WEINSTEIN: Same objection; outside	14:03:17
22	the scope.	14:03:19

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 70

1 You can answer. 14:03:20

2 THE WITNESS: What I'm referring to 14:03:23

3 memorization is the project that was checking if, 14:03:25

4 given a small subsequence of the document, how does 14:03:30

5 the generation of a particular model match or does not 14:03:34

6 match the preceding subsequence, usually limited to 14:03:42

7 50 tokens or so. 14:03:48

8 BY MS. POUEYMIROU: 14:03:49

9 Q. And to prevent memorization in the model, 14:03:49

10 does that involve training the model? 14:03:52

11 MR. WEINSTEIN: Object to form. 14:04:01

12 THE WITNESS: I don't really understand 14:04:03

13 the question. 14:04:07

14 BY MS. POUEYMIROU: 14:04:07

15 Q. My question is that in order to prevent 14:04:08

16 memorization, does that involve training? Is the 14:04:11

17 model trained to not regurgitate memorized text, or 14:04:22

18 does Meta achieve that in a different way? 14:04:27

19 MR. WEINSTEIN: Object to form, compound. 14:04:30

20 THE WITNESS: I would say that memorize -- 14:04:42

21 like, the project that I'm referring to which measures 14:04:47

22 a specific metric of how many tokens does the model 14:04:49

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 73

1	to creating that smaller set?	14:08:36
2	A. There should be in the code base.	14:08:40
3	Q. And where would that be stored? In which	14:08:45
4	source code repository would that be stored in?	14:08:48
5	A. I think -- sorry. Most likely, it	14:08:57
6	should be in XLFormers --	14:09:01
7	Q. Okay.	14:09:06
8	A. -- but I can, yeah, check if that is	14:09:07
9	needed.	14:09:10
10	Q. Okay. Has LibGen been used for ablation	14:09:10
11	studies?	14:09:21
12	MR. WEINSTEIN: Object to form.	14:09:21
13	BY MS. POUYMIROU:	14:09:21
14	Q. You can answer.	14:09:22
15	A. Could you specify which ablation studies	14:09:23
16	or, like, in general?	14:09:26
17	Q. For any ablation studies. Have you used	14:09:27
18	LibGen for any ablation studies?	14:09:30
19	MR. WEINSTEIN: Object to form, scope.	14:09:33
20	You can answer.	14:09:35
21	THE WITNESS: For some of the ablation	14:09:38
22	studies, LibGen was used, yes.	14:09:41

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 74

1	BY MS. POUEYMIROU:	14:09:43
2	Q. And were those ablation studies in the	14:09:43
3	pretraining part of a life cycle?	14:09:45
4	A. Yes.	14:10:00
5	Q. And were there ablation studies using	14:10:01
6	LibGen in the post-training part of the life cycle?	14:10:04
7	A. Yes.	14:10:08
8	Q. Okay. And has LibGen been used for	14:10:08
9	benchmarking?	14:10:11
10	A. So, again, if you could specify what do	14:10:12
11	you mean under this?	14:10:15
12	Q. Okay. We'll look at a document.	14:10:16
13	Has LibGen been used to determine whether	14:10:19
14	to license with different content providers?	14:10:21
15	MR. WEINSTEIN: Object to form, outside	14:10:25
16	the scope.	14:10:28
17	MS. POUEYMIROU: It's a use of LibGen.	14:10:28
18	MR. WEINSTEIN: I -- hard to believe you	14:10:31
19	could have it with a straight face.	14:10:34
20	You can answer that question if you can.	14:10:36
21	Object to form, outside the scope.	14:10:38
22	THE WITNESS: I do not know how LibGen was	14:10:47

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 80

1	definition here so it's difficult for me to assess	14:19:47
2	whether or not it is kind of true.	14:19:51
3	BY MS. POUEYMIROU:	14:19:55
4	Q. Okay. Is it fair to say that however Meta	14:19:55
5	defines "copyright books," it is testing memorization	14:19:58
6	with copyright books?	14:20:01
7	MR. WEINSTEIN: Object to form, outside	14:20:04
8	the scope.	14:20:09
9	THE WITNESS: Based on this document,	14:20:09
10	probably, yes.	14:20:11
11	MS. POUEYMIROU: Okay. We'll mark this	14:20:18
12	as 821.	14:20:19
13	(Plaintiffs' Exhibit 821 marked for	14:20:24
14	identification.)	14:20:39
15	BY MS. POUEYMIROU:	14:20:39
16	Q. So earlier today, Mr. Bashlykov, I asked	14:20:39
17	you about MMLibU, which you said you didn't know what	14:20:41
18	I was talking about.	14:20:45
19	A. Yeah.	14:20:46
20	Q. What is this document called?	14:20:46
21	A. "MMLibU - new eval for Datasets."	14:20:49
22	Q. And who is the author?	14:20:56

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 136

1 Q. But your testimony today is "parts of." 16:25:20

2 Is your testimony today that with respect to LLaMAs 1 16:25:24

3 and 2, that Meta used the Books3 dataset, as opposed 16:25:28

4 to a part of it? 16:25:31

5 A. Could you repeat the question, please? 16:25:44

6 Q. Yes. So you said I could look at LLaMA 1, 16:25:46

7 the LLaMA 1 paper. 16:25:49

8 A. Correct. 16:25:50

9 Q. The LLaMA 1 paper says that LLaMA 1 was 16:25:50

10 trained on Books3 and it says that LLaMA 1 was trained 16:25:53

11 on Gutenberg. 16:25:56

12 My question for you is: The LLaMA 1 paper 16:25:58

13 does not say it was trained on parts of Books3 or 16:26:00

14 parts of Gutenberg. Is it your testimony on behalf of 16:26:03

15 Meta that LLaMA 1 was trained on all of Books3? 16:26:07

16 A. So I think I would answer in the following 16:26:29

17 way: that for LLaMA 1, LLaMA 1 was trained on the 16:26:32

18 dataset names Books3 and Gutenberg that Meta has 16:26:40

19 provided on one of the hard drives, I believe. 16:26:47

20 Q. Okay. And so are you saying that 16:26:49

21 everything in -- if we can pull that out, in fact. 16:26:53

22 Everything in -- so Meta_Kadrey_Data_001, Meta 16:27:00

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 137

1	produced B3G there, and B3G is Books3, right? Books3	16:27:10
2	plus Gutenberg, correct?	16:27:16
3	A. Correct.	16:27:17
4	Q. So your testimony is that the data on this	16:27:17
5	hard drive, with respect to B3G, that LLaMA 1 and	16:27:20
6	LLaMA 2 trained on the entirety of that data?	16:27:29
7	A. As I was mentioning before, I don't think	16:27:39
8	there is, like, a bulletproof method to say that all	16:27:42
9	of the documents that were used because, as I was	16:27:47
10	saying, there could be -- like, when we are using	16:27:49
11	dataloader for the model, when we train the model with	16:27:52
12	the dataloader, usually it does some random sample and	16:27:55
13	doesn't go line by line for each file. So it does	16:28:00
14	some random selection of documents to feed in the	16:28:03
15	model.	16:28:09
16	That's why I'm saying that I can't be a	16:28:11
17	hundred percent sure or, like, confident that all,	16:28:14
18	exactly 100 percent of the documents that were	16:28:19
19	provided to you were used for -- during the training	16:28:23
20	process.	16:28:27
21	Q. So, in your experience, has the dataloader	16:28:28
22	missed data in a dataset?	16:28:33

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Nikolay Bashlykov 30(b)(6)
Highly Confidential - Attorneys' Eyes Only

Page 155

1 CERTIFICATE OF COURT REPORTER

2 I, LEAH M. WILLERSDORF, Registered Merit
3 Reporter, Certified Realtime Reporter, Fellow of the
4 British Institute of Verbatim Reporters, Qualified
5 Realtime Reporter Level 2, and Certified LiveNote
6 Reporter, do hereby certify that:

7 NIKOLAY BASHLYKOV appeared before me on Friday,
8 December 06, 2024, was sworn by me and was thereupon
9 examined by counsel; that the testimony of said
10 witness was taken by me stenographically; that the
11 foregoing is a true and accurate record to the best of
12 my knowledge, skill and ability; that I am neither a
13 relative nor employee of any party to the action in
14 which this testimony was taken; that I am neither
15 relative nor employee of any attorney or counsel
16 employed by any party thereto; and, further, I am not
17 financially or otherwise interested in the outcome of
18 the action.

19

20



21 LEAH M. WILLERSDORF

RMR, CRR, FBIVR, ACR, QRR2, CLR

22

(December 08, 2024)